

# Modelos Aleatorios de Formación de Redes<sup>1</sup>

Alvaro J. Riascos Villegas

Agosto, 2023

---

<sup>1</sup>Basado en Balakrishnan, R y Ranganathan, K (2000). A Textbook of Graph Theory y Jackson, M (2008). Social and Economic Networks. Princeton University Press. Todas las figuras Jackson (2008).

# Contenido

- 1 **Introducción**
- 2 Modelos Estáticos
- 3 Modelos Dinámicos
- 4 Predicción y Prevención de Agentes Dominantes en Redes Complejas

# Modelos de Formación de Redes

- Modelos estáticos: Se refiere a modelos en el que todos los enlaces se establecen al mismo tiempo de forma aleatoria.
- Alternativamente, en los modelos dinámicos existe una red inicial que crece de forma aleatoria y la probabilidad de formación de nuevos enlaces depende del estado actual de la red.
- En los modelos de formación estratégica, la red es el resultado de interacción de agentes con intereses individuales sobre la forma (i.e., topología) de la red.

Estos modelos se utilizan principalmente para:

- Identificar patrones en la formación de redes que guíen la construcción de modelos estructurales (e.g, modelos de comportamiento).
- Entender la forma como se difunden características de los agentes (e.g., información, enfermedades).
- Sirven como benchmark (i.e., referencia).

# Contenido

- 1 Introducción
- 2 Modelos Estáticos**
- 3 Modelos Dinámicos
- 4 Predicción y Prevención de Agentes Dominantes en Redes Complejas

# Modelo de Erdos - Renyi: Modelo Binomial o de Poisson

- Modelo de Erdos - Renyi 1960.
- Sea  $N = \{1, \dots, n\}$  un conjunto de nodos (vertices).
- Suponga que la probabilidad de que se forme un enlace entre  $i, j$  es  $p$  y que la formación de enlaces es independiente.
- En la práctica para simular una red con este mecanismo aleatorio: enumere todas las parejas posibles (sin importar el orden) y después recorra la lista de parejas haciendo enlaces de forma independiente con probabilidad  $p$ .

## Example

Si  $n = 3$  la probabilidad de que se forme una red completa (i.e., todos los nodos enlazados con todos) es  $p^3$  y hay una sola red completa.

La probabilidad de que se forme una red específica con dos enlaces es  $p^2(1 - p)$  y hay tres redes de esta forma.

## Example

Con  $n$  nodos la probabilidad de que se forme una red específica de  $m$  enlaces es:

$$p^m(1 - p)^{\binom{n(n-1)}{2} - m} \quad (1)$$

Obsérvese que estas probabilidades son de redes donde importa la identidad de los nodos. Por ejemplo, la probabilidad de que se forme *alguna* red de tres nodos con dos enlaces es:  $3p^2(1 - p)$ .

- Una forma alternativa de describir este tipo de redes es la siguiente. Sea  $G(n)$  el conjunto de todas las redes con  $n$  nodos y supongamos que cada red con  $m$  enlaces se selecciona con probabilidad:

$$p^m(1 - p)^{\binom{n(n-1)}{2} - m} \quad (2)$$

- Llamamos este modelo de formación de redes el modelo  $G(n, p)$ .



- Sea  $G(n)$  el conjunto de todas las redes con  $n$  vertices. El número de elementos en  $G(n)$  es  $2^{\binom{n}{2}}$ . Elegir de  $G(n)$  cada red con la misma probabilidad.

# Formación aleatoria de redes

- Retomemos nuestro modelo de formación aleatoria de Erdos y Reny.
- Calculemos algunas estadísticas descriptivas. El grado de un nodo es el número de nodos con los cuales se tiene un enlace.
- La probabilidad de que un nodo  $i$  tenga exactamente  $d$  enlaces es:

$$\binom{n-1}{d} p^d (1-p)^{n-1-d} \quad (3)$$

- Cuando  $n$  es grande y  $p$  es pequeño esta expresión se puede aproximar por una distribución de Poisson (en ocasiones el modelo binomial independiente lleva este nombre - modelo de Poisson).

$$\frac{e^{(n-1)p} ((n-1)p)^d}{d!} \quad (4)$$

- Para ver esto, cuando  $n$  es grande y  $p$  es pequeño, obsérvese que:

$$(1 - p)^{n-1-d} \approx (1 - p)^{n-1} \approx e^{-(n-1)p}$$

por otro lado:

$$\binom{n-1}{d} \approx \frac{(n-1)^d}{d!} \quad (5)$$

# Ejemplos de Redes: Formación aleatoria de redes

- Las siguientes figuras muestran una red generada usando el modelo binomial ( $n = 50, p = 0,02$ , lo cual implica que el valor esperado del grado de un nodo es 1).
- Características sobresalientes de grafos generados de esta forma: la probabilidad de ciclos es baja y existe una componente conexas grande.

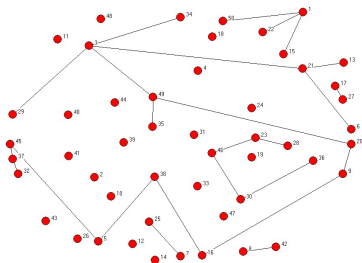


Figure 1.4: A Randomly Generated Network with Probability .02 on each Link

# Ejemplos de Redes: Formación aleatoria de redes

- El siguiente gráfico compara la distribución muestral con la aproximación de Poisson para la distribución binomial.

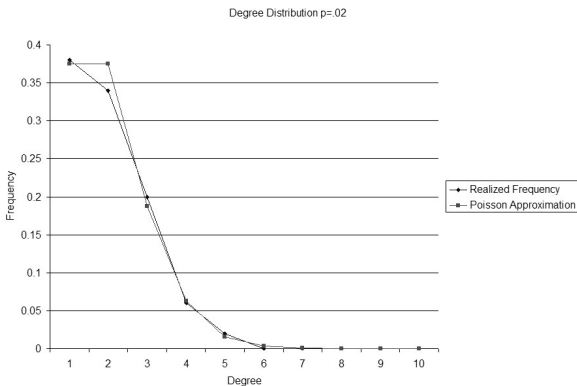


Figure 1.5: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .02 on each Link

# Ejemplos de Redes: Formación aleatoria de redes

- Al aumentar la probabilidad de enlaces disminuyen las componentes conexas (maximales), aumenta la componente dominante y empeora la aproximación de Poisson al grado del grafo.

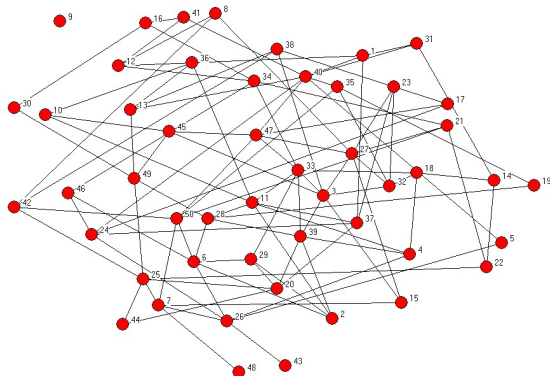


Figure 1.6: A Randomly Generated Network with Probability .08 of each Link

# Ejemplos de Redes: Formación aleatoria de redes

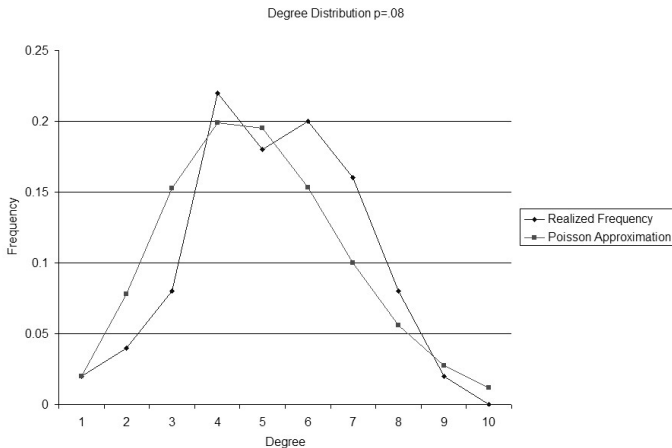


Figure 1.7: Frequency Distribution of a Randomly Generated Network and the Poisson Approximation for a Probability of .08 on each Link

# Ejemplos de Redes: Formación aleatoria de redes

- Un problema con este modelo son los bajos niveles de aglomeración que se obtienen (i.e., índice de clustering bajo).
- En particular, si el grado promedio crece a una tasa menor que el número de nodos, el coeficiente de aglomeración (clustering) tiende a cero.
- El modelo de Watts y Strogatz ataca este problema (pero no el siguiente).
- La distribución del grado tiende a una Poisson y no a una power law, común en redes (el modelo de enlaces preferenciales, preferential attachment, si intenta capturar esto).



## Modelo de configuración

- Tiene como objetivo generar redes aleatorias controlando la distribución del grado.
- Sea  $\{d_1, \dots, d_n\}$  los grados de una red de  $n$  vertices. La sucesión determina la distribución del grado
 
$$P^n(d) = \frac{\text{num}\{i:d_i=d\}}{n}.$$
- Ahora, dados  $n$  nodos queremos construir una red aleatoria con aproximadamente el grado del vector  $\{d_1, \dots, d_n\}$ .
- Considere el siguiente procedimiento. Construir una lista:

$$1, \dots, 1, (d_1 \text{ veces}), 2, \dots, 2, (d_2 \text{ veces}), \dots \quad (6)$$

- Aleatoriamente seleccione dos elementos de la sucesión, cree un enlace entre los nodos a los que corresponde y elimínelos de la lista.

# Modelo de configuración

- Obsérvese que pueden haber loops y múltiples enlaces (multigrafo). Además si el número de nodos no es par, al final puede quedar un enlace sin formarse.
- El modelo es diseñado para reproducir aproximadamente el grado del vector dado. Sin embargo, nada garantiza que se puedan reproducir otra propiedades.
- Por ejemplo, usualmente no genera patrones de cohesión como los observados en redes reales.

# Aplicación: Contagio y difusión enfermedades

- Variación del modelo Reed-Frost en literatura de epidemiología.
- Población inicial de  $n$  individuos.
- Con probabilidad  $p$  un individuo interactúa de forma independiente con otro individuo, algunos de los cuales pueden ser inmunes.
- Con probabilidad  $\pi$  un individuo es inmune.
- Inicialmente un individuo tiene una infección.

La difusión de la enfermedad puede modelarse así:

- Generar una red aleatoria de  $n$  nodos con el modelo Binomial con probabilidad  $p$ .
- Elegir aleatoriamente con la misma probabilidad  $n\pi$  nodos y eliminarlos.
- En la red resultante identificar la componente con la persona inicialmente infectada.

Alternativamente:

- Generar una red aleatoria de  $n(1 - \pi)$  nodos con el modelo Binomial con probabilidad  $p$ .
- Elegir un nodo de forma aleatoria (la persona inicialmente infectada) y estudiar la componente conexa en la que se encuentra.
- En el modelo de Poisson se sabe que el surgimiento de una componente gigante lo determina  $p(1 - \pi)n = 1$  (i.e., transición de fase). Si  $p(1 - \pi)n > 1$  con probabilidad mayor que cero se infecta un número considerable de la población.
- La siguiente figura muestra la fracción esperada de la población en la componente gigante (i.e., personas susceptibles) como función del factor que determina la transición de fase.

# Contenido

- 1 Introducción
- 2 Modelos Estáticos
- 3 Modelos Dinámicos**
- 4 Predicción y Prevención de Agentes Dominantes en Redes Complejas

## Poisson Dinámico: Grado Esperado Exponencial

- Supongamos que cada cierto tiempo se crea un nodo y este forma enlaces con los anteriores de la siguiente forma.
- Sea  $i \in \{0, 1, 2, 3, \dots\}$  un nodo que nace en el momento  $i$ .
- Supongamos que comenzamos con una red completa de  $m$  nodos:  $\{0, 1, 2, 3, \dots, m - 1\}$  y el primer nodo que nace es  $m + 1$ .
- Cada que nace un nodo, se eligen aleatoriamente  $m$  de los nodos existentes para formar enlaces con ellos.
- El valor esperado del grado de un nodo que nació en  $i \geq m$  en un periodo  $t > i$  es:

$$E[d_i(t)] = m + \frac{m}{i+1} + \frac{m}{i+2} + \dots + \frac{m}{t}$$

para  $t$  grande  $E[d_i(t)] \approx m(1 + \log(\frac{t}{i}))$

- Luego, la proporción de nodos con valor esperado del grado menor de  $d$  hasta el momento  $t$  es:  $F_t(d) = 1 - \exp(-\frac{d-m}{m})$ .
- En lo que sigue abusando de la notación denotamos  $E[d_i(t)]$  por  $d_i(t)$ .



# Poisson Dinámico: Grado Esperado Exponencial - Aproximación de Campo Medio

- Sabemos que  $d_i(i) = m$  y postulamos que hasta el momento  $t$  es:  $\frac{dd_i(t)}{dt} = \frac{m}{t}$ . Es decir, los cambios en el grado de  $i$  en  $t$  dependen de la probabilidad de hacer un enlace con un nodo recién nacido (i.e., que debe repartir su enlaces con  $m$  de  $t$  nodos).
- La clave para usar la aproximación de campo medio es poder determinar la probabilidad de que un nodo establezca un nuevo enlace con un nodo recién nacido.
- Se puede demostrar que la distribución de nodos con grado menor que  $d$  en  $t$  es igual a la fórmula anterior.

# Preferential Attachment

- Cada nuevo nodo, en vez de seleccionar de forma uniforme con cuales de  $m$  nodos hacer un enlace, es mas probable hacerlo con aquellos que ya tiene bastantes enlaces (i.e., pensar en como crece una red de citas).
- La probabilidad de que un nodo existente  $i$  establezca un enlace con un recién nacido en  $t$  se postula como:

$$m \frac{d_i(t)}{\sum_{j=1}^t d_j(t)} \quad (7)$$

pero el número total de enlaces en la red en  $t$  es  $2m$ , luego el denominador es  $2tm$ .

- Se sigue que la probabilidad de que un nodo existente  $i$  establezca un enlace con un recién nacido en  $t$  es:

$$\frac{d_i(t)}{2t} \quad (8)$$

- Luego la aproximación de campo medio es:

$$\frac{dd_i(t)}{dt} = \frac{d_i(t)}{2t} \quad (9)$$

- Luego la fracción de nodos que tienen un grado (esperado) menor que  $d$  es:

$$F_t(d) = 1 - m^2 d^{-2} \quad (10)$$

# Contenido

- 1 Introducción
- 2 Modelos Estáticos
- 3 Modelos Dinámicos
- 4 Predicción y Prevención de Agentes Dominantes en Redes Complejas

## Preferential Attachment y Fitness

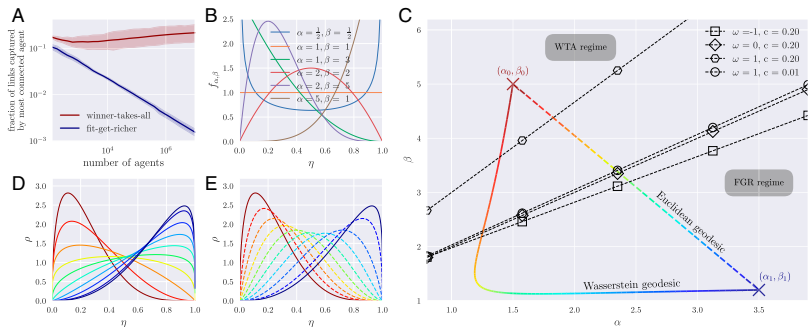
- La aproximación de campo medio es:

$$\frac{dd_i(t)}{dt} = m \frac{\eta_i d_i(t)}{S(t)} - \frac{c}{(1-c)} \frac{\bar{\eta}^w}{\eta_i^w} \frac{d_i}{t}$$

donde  $\eta_i \in (0, 1)$  es el fitness (aptitud) de  $i$ ,  $\rho$  es una distribución del fitness con la que se calculan los promedios arriba,  $S(t)$  es un factor de normalización y las demás son constantes.

- Aquí  $\frac{\bar{\eta}^w}{\eta_i^w}$  captura la posibilidad de ser eliminado de la red (muerte, bancarrota, etc) y  $\frac{d_i}{t}$  la probabilidad de que  $i$  tenga un enlace con el nodo eliminado.
- El diagrama de fase de la siguiente figura depende de  $I(\rho, w, c)$ . Es decir de la distribución completa del fitness.

# Preferential Attachment y Fitness



**Fig. 1.** (A) Fraction of links connected to the most dominant agent  $n_{DK}$  in the system, as a function of the total number of agents. The mean values and  $\pm$  one SD bands are obtained over a sample of 100 realizations. We have fixed  $c = 1\%$  and  $\omega = 0$ . Two different fitness distributions are used, one in the WTA and one in the FGR regime (red and blue crosses in C). (B) Beta distribution density for different parameters  $\alpha$  and  $\beta$  used to parametrize the class of fitness distributions on  $[0, 1]$ . As  $\alpha$  and  $\beta$  vary, all typically relevant shapes of the fitness distribution are sampled. The WTA regime is characterized by the bulk of the probability mass being concentrated on agents of relatively low fitness, with only few fit agents that end up dominating the system (e.g.,  $\alpha = 2, \beta = 5$ ). (C) Phase diagram showing the domains of existence for the FGR and the WTA regimes, for different preferential deletion parameters  $c$  and  $\omega$  in the space of the parameters  $(\alpha, \beta)$  of the beta-fitness distribution. The WTA regime corresponds to values where  $I^* < 1$  (Eq. 2). This WTA regimes lies above a line parameterized by  $(c, \omega)$ . (D) Geodesic with respect to Wasserstein-2 metric, that is, cost minimizing interpolation from a beta distribution in the WTA regime to a beta distribution in the FGR regime. The (approximate) path in the space of beta distributions is shown in C. (E) Distributions along the shortest path with respect to the Euclidean distance in the  $(\alpha, \beta)$  plane (cf. dashed line in panel C). This intervention policy is more expensive than the one along the Wasserstein geodesic.